

**Title Page:****Genome-wide association for major depression through age at onset stratification**

Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

**Authors:**

Robert A. Power<sup>1\*</sup>, Katherine E. Tansey<sup>2\*</sup>, Henriette Nørmølle Buttenschøn<sup>3,4</sup>, Sarah Cohen-Woods<sup>5</sup>, Tim Bigdeli<sup>6</sup>, Lynsey S. Hall<sup>7</sup>, Zoltán Kutalik<sup>8</sup>, S. Hong Lee<sup>9,10</sup>, Stephan Ripke<sup>11,12,13</sup>, Stacy Steinberg<sup>14</sup>, Alexander Teumer<sup>15</sup>, Alexander Viktorin<sup>16</sup>, Naomi R. Wray<sup>9</sup>, Volker Arolt<sup>17</sup>, Bernard T. Baune<sup>5</sup>, Dorret I. Boomsma<sup>18</sup>, Anders D. Børglum<sup>3,19</sup>, Enda M. Byrne<sup>9</sup>, Enrique Castelao<sup>20</sup>, Nick Craddock<sup>2</sup>, Ian Craig<sup>1</sup>, Udo Dannlowski<sup>17,21</sup>, Ian J. Deary<sup>22,23</sup>, Franziska Degenhardt<sup>24,25</sup>, Andreas J. Forstner<sup>24,25</sup>, Scott D. Gordon<sup>26</sup>, Hans J. Grabe<sup>27</sup>, Jakob Grove<sup>3,19</sup>, Steven P. Hamilton<sup>28</sup>, Caroline Hayward<sup>29</sup>, Andrew C. Heath<sup>30</sup>, Lynne J. Hocking<sup>31</sup>, Georg Homuth<sup>32</sup>, Jouke J. Hottenga<sup>18</sup>, Stefan Kloiber<sup>33</sup>, Jesper Krogh<sup>34</sup>, Mikael Landén<sup>16,35</sup>, Maren Lang<sup>36</sup>, Douglas F. Levinson<sup>37</sup>, Paul Lichtenstein<sup>16</sup>, Susanne Lucae<sup>33</sup>, Donald J. MacIntyre<sup>7</sup>, Pamela Madden<sup>30</sup>, Patrik K.E. Magnusson<sup>16</sup>, Nicholas G. Martin<sup>26</sup>, Andrew M. McIntosh<sup>7,22</sup>, Christel M. Middeldorp<sup>18</sup>, Yuri Milaneschi<sup>38</sup>, Grant W. Montgomery<sup>26</sup>, Ole Mors<sup>3,39</sup>, Bertram Müller-Myhsok<sup>33,40,41</sup>, Dale R. Nyholt<sup>42</sup>, Hogni Oskarsson<sup>43</sup>, Michael J. Owen<sup>2</sup>, Sandosh Padmanabhan<sup>44</sup>, Brenda W.J.H. Penninx<sup>38</sup>, Michele L. Pergadia<sup>45</sup>, David J. Porteous<sup>29</sup>, James B. Potash<sup>46</sup>, Martin Preisig<sup>20</sup>, Margarita Rivera<sup>1,47</sup>, Jianxin Shi<sup>48</sup>, Stanley I. Shyn<sup>49</sup>, Engilbert Sigurdsson<sup>50,51</sup>, Johannes H. Smit<sup>38</sup>, Blair H. Smith<sup>52</sup>, Hreinn Stefansson<sup>14</sup>, Kari Stefansson<sup>14</sup>, Jana Strohmaier<sup>36</sup>, Patrick F. Sullivan<sup>16,53</sup>, Pippa Thomson<sup>22,29</sup>, Thorgeir E. Thorgeirsson<sup>14</sup>, Sandra Van der Auwera<sup>27</sup>, Myrna M. Weissman<sup>54</sup>, CONVERGE consortium<sup>55</sup>, CARDIoGRAM Consortium<sup>56</sup>, GERAD1 Consortium<sup>57</sup>, Gerome Breen<sup>1</sup>, and Cathryn M. Lewis<sup>1</sup>

**Affiliations:**

1 Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK

2 MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK

3 The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus University, Aarhus, Denmark

4 Translational Neuropsychiatry Unit, Department of Clinical Medicine, Aarhus University, Aarhus, Denmark

5 Discipline of Psychiatry, School of Medicine, University of Adelaide, Adelaide, Australia

6 Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, USA

7 Division of Psychiatry, University of Edinburgh, Edinburgh, UK

8 Institute of Social and Preventive Medicine (IUMSP), Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland

9 The University of Queensland, Queensland Brain Institute, Brisbane, Queensland, Australia

10 School of Environmental and Rural Science, University of New England, Armidale, New South Wales, Australia

11 Stanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, MA, USA

12 Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

13 Department of Psychiatry and Psychotherapy, Charité, Campus Mitte, 10117 Berlin, Germany

14 deCODE Genetics, 101 Reykjavik, Iceland

15 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

16 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

17 Department of Psychiatry and Psychotherapy University of Muenster, Muenster, Germany

18 Department of Biological Psychology, EMGO Institute for Health and Care Research and Neuroscience Campus Amsterdam, Vrije Universiteit, Amsterdam, The Netherlands

19 Department of Biomedicine and Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark

20 Department of Psychiatry, Lausanne University Hospital, Lausanne, Switzerland

21 Department of Psychiatry, University of Marburg, Marburg, Germany

22 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

23 Department of Psychology, University of Edinburgh, Edinburgh, UK

24 Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany

25 Department of Genomics, Life & Brain Center, University of Bonn, D-53127 Bonn, Germany

26 QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

27 Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany

28 Department of Psychiatry, Kaiser Permanente San Francisco Medical Center, San Francisco, CA, USA

29 Medical Genetics Section, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine,  
University of Edinburgh, Edinburgh, UK

30 Department of Psychiatry, Washington University St. Louis, St. Louis, MO, USA

31 Division of Applied Health Sciences, University of Aberdeen, Aberdeen, UK

32 Interfaculty Institute for Genetics and Functional Genomics, University of Greifswald, Germany

33 Max Planck Institute of Psychiatry, Munich, Germany

34 Mental Health Center Copenhagen, Mental Health Services in Capital Region, University of Copenhagen, Denmark

35 Institute of Neuroscience and Physiology, University of Gothenburg, Gothenburg, Sweden

36 Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim / Heidelberg University, D-68159 Mannheim, Germany

37 Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

38 Department of Psychiatry, EMGO Institute for Health and Care Research and Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands

39 Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark

40 Munich Cluster for Systems Neurology (SyNergy), Munich, 81377, Germany

41 University of Liverpool, Institute of Translational Medicine, Liverpool, L69 3BX, UK

42 Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia

43 Therapiea, 101 Reykjavik, Iceland

44 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

45 Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL, USA

46 Department of Psychiatry, University of Iowa Carver College of Medicine, Iowa City, IA, USA

47 CIBERSAM-Universidad de Granada e Instituto de Investigación Biosanitaria ibs.GRANADA. Hospitales Universitarios de Granada/Universidad de Granada, Granada, Spain

48 Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA

49 Department of Psychiatry, Group Health, Seattle, WA, USA

50 University of Iceland

51 Department of Psychiatry, Landspítali University Hospital, Reykjavik, Iceland

52 Division of Population Health Sciences, University of Dundee, Dundee, UK

53 Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

54 College of Physicians and Surgeons and the Mailman School of Public Health, Columbia University and New York State Psychiatric Institute, New York, NY, USA

55 see supplementary materials for full list of CARDIoGRAM authors 56 see supplementary materials for full list of CARDIoGRAM authors

57 Data used in the preparation of this article were obtained from the Genetic and Environmental Risk for Alzheimer's disease (GERAD1) Consortium. As such, the investigators within the GERAD1 consortia contributed to the design and implementation of GERAD1 and/or provided Data but did not participate in analysis or writing of this report. A full list of GERAD1 investigators shall be included in either supplementary content or acknowledgements.

\*Both authors contributed equally

**Corresponding Author:**

Cathryn M. Lewis

MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London

De Crespigny Park

London SE5 8AF

United Kingdom

Phone: +44-20 7848 0873

Fax: +44-20 7848 0866

Email: [cathryn.lewis@kcl.ac.uk](mailto:cathryn.lewis@kcl.ac.uk)

**Abstract Word Count:** 222

**Article Word count:** 3,783

**Number of Figures:** 3

**Number of Tables:** 2

**Number of Supplemental Information:** 1

**Running Title:** Genomic analysis of age at onset in depression

**Keywords:** Major Depressive Disorder; GWAS; Age at Onset; Polygenic Scoring

**Abstract**

**Background:** Major depressive disorder (MDD) is a disabling mood disorder and, despite a known heritable component, a large meta-analysis of GWAS revealed no replicable genetic risk variants. Given prior evidence of heterogeneity by age-at-onset (AAO) in MDD, we tested whether genome-wide significant risk variants for MDD could be identified in cases subdivided by AAO.

**Method:** Discovery case-control GWASs were performed where cases were stratified using increasing/decreasing AAO-cutoffs; significant SNPs were tested in nine independent replication samples, giving a total sample of 22,158 cases and 133,749 controls for sub-setting. Polygenic score analysis was used to examine if differences in shared genetic risk exists between earlier and adult onset MDD with commonly co-morbid disorders of schizophrenia, bipolar disorder, Alzheimer's disease, and coronary artery disease.

**Results:** We identify one replicated genome-wide significant locus associated with adult-onset (>27 years) MDD (rs7647854, OR=1.16, 95%CI=1.11-1.21,  $p=5.2 \times 10^{-11}$ ). Using polygenic score analyses, we show that earlier-onset MDD is genetically more similar to schizophrenia and bipolar disorder than adult-onset.

**Conclusions:** We demonstrate that using additional phenotype data previously collected by genetic studies to tackle phenotypic heterogeneity in MDD can successfully lead to the discovery of genetic risk factor despite reduced sample size. Furthermore, our results suggest that the genetic susceptibility to MDD differs between adult- and earlier-onset MDD, with earlier-onset cases having a greater genetic overlap with schizophrenia and bipolar disorder.

## Introduction

Major depressive disorder (MDD) is a highly prevalent and heterogeneous disorder (1). With most individuals experiencing recurrent episodes throughout life (2), MDD is now the second leading cause of disability worldwide (3). MDD is defined by low mood and energy, inability to experience enjoyment, changes to eating and sleep patterns, feelings of guilt or worthlessness, and suicidal thoughts (4). Along with excess mortality and increased risk of suicide (5), MDD is associated with worse clinical outcomes when co-morbid with health problems such as cardiovascular disease and cancer (6; 7). While the heritability is estimated at 31-42% (8), the causal variants remain elusive: a recent large mega-analysis with over 9000 MDD cases failed to identify any replicable associations (9), despite successes in similarly sized studies of schizophrenia and bipolar disorder (10; 11). This lack of biological markers may be one of the causes for the well established underfunding of research into MDD relative to its economic and health burden (12), and the reported stigmatization of sufferers (13).

However, several differences exist between MDD and other psychiatric disorders where replicable genetic associations have been identified, including higher prevalence, greater diagnostic uncertainty, lower heritability, and, crucially, increased heterogeneity. One known source of heterogeneity that may contribute substantially is age at onset (AAO). Onset can occur at any stage of life, yet many factors associated with MDD are either age specific or restricted. These include biological events such as puberty, menopause and dementia, and environmental risk factors including childhood maltreatment, childbirth, and divorce. Earlier onset is associated with increased risk in first-degree relatives and higher heritability (14-18). Considerable differences in the transmission of early versus late onset MDD have also been reported (19; 20), with some studies suggesting the effects of novel genetic risk factors for MDD appearing later in life (21; 22).

Here we build on the previous mega-analysis of Major Depressive Disorder Working Group of the PGC (PGC-MDD), using AAO to stratify cases within a sample of 8920 cases and 9521 controls, with the goal of reducing heterogeneity. For each of the nine PGC-MDD samples (23-30), cases were ordered by AAO within study and divided into eight groups (octiles). Genome-wide association analysis of cases in these octiles was performed systematically against controls for 1 235 109 autosomal



single nucleotide polymorphisms (SNPs). We examined three analytic strategies: (a) genetic variants specific to early-onset MDD; (b) risk variants specific to late-onset; and (c) restricting to the intermediate four octiles excluding the 25% of cases at either extreme of AAO to test for potential heterogeneity introduced from very early or very late onset. Significantly associated SNPs were taken forward for replication in nine studies comprising 13 238 cases and 124 230 controls. We also examined the differences between early and late onset MDD in their shared heritability with commonly co-morbid disorders of schizophrenia, bipolar disorder, Alzheimer's disease, and coronary artery disease, in order to identify differences in genetic etiologies across onset groups.

## **Methods and Materials:**

### ***Description of samples***

Full details of the studies that form the PGC-MDD are given in the supplementary materials of the original data analysis (9). Briefly, these nine MDD studies (23-30) conducted genome-wide genotyping on individual subjects of European ancestry. Cases were required to have diagnoses of DSM-IV lifetime MDD established using structured diagnostic instruments from direct interviews by trained interviewers or clinician-administered DSM-IV checklists. Two studies required recurrent MDD and one recurrent, early-onset MDD. Studies ascertained cases mostly from clinical sources, and controls were largely randomly selected from the population and screened for lifetime history of MDD. This led to a total of 9238 cases and 9521 controls with genotype information.

### ***AAO phenotype***

Age at onset (AAO) was defined as the age at which individuals first had symptoms that met the criteria of MDD, and was self-reported in all studies. Of the original 9238 cases included in the sample, 8920 (95.6%) had a reported AAO. Cases reporting an AAO older than the recorded age at interview were removed from the analysis (n=17). Within each study, cases were ordered by AAO, and then divided into octiles, giving approximately 1000 cases per octile. Octiles were defined within study to account for differences in case ascertainment. We noted a wide range in AAO between studies ascertaining recurrent

depression using the same instrument, indicating that the precise setting (study, clinic, country) was important; we therefore chose to order cases by AAO within study, rather than across studies, or by absolute AAO cut-offs. This strategy will identify genetic variants that were specific to early or late onset, relative to the mean AAO of the recruited cases. For secondary analysis of sex specific effects, octiles were additionally defined within only males and females for analysis of sex-specific effects, and for recurrent depression that has a higher heritability. These octiles will be referred to as O1-8, with O1 representing the earliest onset octile, and O8 representing the latest onset octile. The GenRED study only recruited MDD cases with an age at onset below 31 years (25). We compared the distribution of AAO in GenRED with other similar studies (STAR\*D, RADIANT-UK) and estimated that GenRED recruited the youngest 62% of all possible MDD cases, with older cases absent from the study (**Supplementary Figure 1**). GenRED cases were ordered by AAO, and assigned to the appropriate octiles O1-5, with no cases present in O6-8.

### ***Quality Control***

Genotyping was described in the supplementary materials in the original analysis.<sup>11</sup> All samples were genotyped with single nucleotide polymorphism (SNP) arrays of at least 200 000 SNPs. SNPs were removed for missingness >0.02, case–control difference in SNP missingness >0.02, SNP frequency difference from HapMap3 >0.15, or deviation from Hardy–Weinberg equilibrium in controls (exact  $p < 1.0 \times 10^{-6}$ ). Subjects were removed for excessive missingness (>0.02), identical or closely related to any subject in any sample ( $\pi$ -hat > 0.2 based on common autosomal SNPs), or if there was evidence for diverging ancestry. Ancestry was estimated using multidimensional scaling applied to 8 549 SNPs directly genotyped in all samples and in approximate linkage equilibrium. Imputation was performed using Beagle 3.0.440 (31) with the CEU HapMap3 data (32) to impute 1 235 109 autosomal SNP allele dosages. The first 20 ancestry–informative principal components were included as covariates, along with an indicator for each study.

### ***Genome-wide association analysis***

Genome-wide association analysis was performed in PLINK using logistic regression to test the association between case control phenotype and imputed SNP dosages under an additive model (33). Genotyping coordinates are given in NCBI Build

36/UCSC hg18. Quality control was conducted separately for each sample. To test for SNPs associated with distinct aspects of MDD based on AAO, we performed three hypothesis-driven analyses (**Supplementary Figure 2**). The first analysis targeted those SNPs associated with early-onset MDD in a series of genome-wide association analyses, initially looking at the earliest onset cases (O1) against all controls, then the combined O1-2 cases against controls, then O1-3, etc. until all cases were included. This approach was based on the Sequential Additions Method (34), developed to take account of a quantitative trait that is measured in cases but not controls and provide an estimate of the best phenotype definition for future studies. The second analysis was similarly performed to examine those SNPs associated with later onset MDD, but reversing the procedure (i.e., first looking at the latest onset cases (O8) against all controls, then O7-8 against controls, then O6-8, etc. until all cases were included). The third analysis tested if the extremes of AAO, both early and late, were introducing heterogeneity to the cases excluding O1-2 and O7-8, leaving those cases with onset within the inter-quartile range of AAO for each study (O3-6). We then tested O3-6 cases against all controls. Each analysis was performed using all cases; male cases only; female cases only; and recurrent cases only. Analyses of all, or almost all, cases (O1-O8, O1-O7, or O2-8) were used to identify SNPs that reached greatest significance without an AAO specific effect, and so were omitted from further analysis as would have been captured in the primary analysis of this data. In total, 52 genome-wide analyses were performed  $(=(6+6+1) \times 4)$ , making the standard multiple testing threshold of  $p < 5.0 \times 10^{-8}$  for genome-wide significance anti-conservative. We applied a Bonferroni correction for 52 analyses to genome-wide significance, which is highly conservative since many analyses were highly correlated. For replication, we selected SNPs with  $p < 9.5 \times 10^{-10}$  in the discovery sample in more than one analysis (either by sex, recurrence or inclusion of octiles); the combination of cases that yielded the greatest significance was chosen as the basis for replication. This analysis strategy uses AAO as a stratifying variable to construct subsets of cases which may be more homogeneous and identify SNPs that are associated with susceptibility to MDD with a restricted AAO; it does not identify SNPs that control MDD AAO, which would require a case-only analysis of AAO as a quantitative trait.

### ***Replication analysis***

Five replication samples used in the primary analysis of this dataset had AAO information available (TwinGene (35), GenREDII/DepGenesNetwork (25), deCODE (9), PsychCoLaus (36), SHIP-LEGEND (37)). The GenPod/NEWMEDS and Harvard

12B2 studies, which appeared in the replication of the primary analysis of this dataset, did not have AAO data available and were not included. Four new replication studies were available ((1) a collection of samples available through the University of Münster (38; 39); (2) a combination of RADIANT cases from Denmark (40), the Danish DEMO and PRISME studies of MDD (41; 42), and a set of Danish population controls; (3) the CONVERGE study of MDD cases and controls recruited in China (18; 43; 44); and (4) the Generation Scotland study which included measures on MDD (45)). These are outlined, alongside the definitions of AAO, in the Supplementary Methods. Due to the early median AAO in the GenREDII/ DepGenesNetwork, an artificial ‘median’ was introduced at age 27, based on the median for the discovery samples. Those SNPs that passed our threshold of  $p < 9.5 \times 10^{-10}$  in the discovery sample were tested for association within these nine replication studies. Due to differences in availability of genome-wide genotype data, each study was genotyped and/or imputed separately. A fixed-effect inverse variance based meta-analysis of the replication studies was performed using METAL (46).

### ***Polygenic analysis***

We also examined the association between early/late onset MDD and polygenic risk scores for other psychiatric disorders, as this might reflect either shared genetic etiology or phenotypic contamination (47). Polygenic risk scores for schizophrenia (9379 cases and 7736 controls) and bipolar disorder (6990 cases and 4820 controls) were created by the PGC, using imputed data and removing overlapping controls, ensuring the datasets were completely independent (10; 11). Alzheimer’s disease polygenic risk scores were obtained from Genetic and Environmental Risk for Alzheimer’s disease (GERAD) Consortium (48) and the coronary artery disease scores from the CARDIoGRAM Consortium (49). The imputed GERAD sample comprised 3177 cases and 7277 controls, and the CARDIoGRAM consortium consisted of 22 233 cases and 64 762 controls. All four disorders were chosen for their previous genetic and epidemiological evidence of overlap with MDD (50-52).

Polygenic risk scores were calculated for MDD cases and controls, summing the number of risk alleles carried, weighted by the natural log of their odds ratio in the original genome-wide association study. Score SNPs with low minor allele frequency (MAF<0.02) or in the major histocompatibility complex (MHC) region were removed, and score datasets were pruned for linkage disequilibrium using the clumping command in PLINK to remove SNPs within 500kb and  $r^2 > 0.25$  with a more

significantly associated SNP. Seven scores were calculated, using a p-value threshold ( $P_T$ ) to restrict to the most significant SNPs in their original genome-wide association analysis ( $P_T < 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$ ). Logistic regression was used to test for association between polygenic risk scores and case control status using four different case-control sets (all MDD cases vs. controls, O1-3 cases vs. controls, O6-8 cases vs. controls, and O1-3 vs. O6-8 cases) with 20 population principal components and study indicators as covariates. We calculated the proportion of variance explained (Nagelkerke's  $R^2$ ) by subtraction of a full model (covariates + RPS) score from a reduced model (covariates only). The GenRED sample was not included in the analysis as they had no cases within the O6-8 analysis; the Bonn-Mannheim study was not included in the analysis of Alzheimer's disease or coronary artery disease due to overlapping controls with the respective consortia.

## Results

### *Summary of AAO*

After quality control the sample consisted of 8920 MDD cases with AAO information and 9519 controls with 1 235 109 SNPs. The median AAO across all cases excluding GenRED was 27 years old (IQR=18-38) and the mean 28.9 (SD 13.64), reflecting the long tail of older onset cases (**Figure 1 and Supplementary Figure 3**). Median AAO by study ranged from 20 - 37 years (16 for GenRED which recruited only recurrent cases with onset no greater than 30), with the German samples having a slightly older onset. Mean AAO was lower for recurrent cases than non-recurrent, 27.1 to 33.0 ( $p < 0.001$ , correcting for study and excluding GenRED). AAO was also older for males with a mean of 29.0 compared to 26.7 in females ( $p < 0.001$ ).

### *Genome-wide association analysis*

Our analysis of both early and late-onset octiles and the intermediate median of AAO excluding extreme-AAO cases led to four tests passing our significance cut off for replication ( $p < 9.5 \times 10^{-10}$ ), all for associations with SNP rs7647854 on chromosome 3. The only genome-wide significant association for this SNP was in the 50% oldest onset cases against all controls (O5-8:  $p = 3.4 \times 10^{-11}$ , **Supplementary Figure 4 & 5**). As a secondary analysis of this SNP, we split cases within our discovery sample into non-overlapping quartiles and analyzed against controls. This showed a strong association in the oldest quartile (O7-8:

$p=9.0 \times 10^{-10}$ , OR=1.37, 95%CI 1.23-1.51; **Figure 2**), a moderate effect in the 3<sup>rd</sup> quartile (O5-6:  $p=2.0 \times 10^{-5}$ , OR=1.23), and no evidence of association in the two youngest quartiles (O1-2:  $p=0.07$ , OR=1.09; O3-4:  $p=0.30$ , OR=1.06). Further, when the inclusion threshold for AAO was increased one year at a time within each study (until only 100 cases remained), we observed a gradual increase in effect size for rs7647854 in all studies as the analysis was restricted to progressively later onset cases (**Supplementary Figure 6**). Analysis by sex or recurrence revealed no additional findings.

### ***Replication analysis***

Given the results from our discovery data set, rs7647854 was taken forward for replication in nine independent studies (**Table 1**). The SNP was tested for association with oldest half of MDD cases (O5-O8), since this analysis attained the smallest p-value in the discovery studies. The SNP was either directly genotyped or imputed at high confidence across all studies (**Supplementary Table 1**). The SNP was significantly associated with MDD in a meta-analysis of these studies, with a p value of  $7.5 \times 10^{-4}$  and an odds ratio of 1.10 (total number of MDD cases = 6,107 and total number of controls 124,230) (**Table 2; Supplementary Figure 5**). Meta-analysis of the combined discovery sample with individual replication studies gave a p-value of  $5.2 \times 10^{-11}$  and an odds ratio of 1.16 (95%CI 1.11-1.21), surpassing genome-wide significance.

### ***Polygenic analysis of co-morbid illnesses***

In the PGC MDD discovery studies, polygenic risk scores for bipolar disorder (BPD) and schizophrenia (SCZ) were significantly associated with early-onset MDD. Restricting to only early-onset cases (O1-3) vs. controls, the amount of phenotypic variability explained (BPD  $R^2=0.41\%$ ,  $p=1.4 \times 10^{-12}$ ; SCZ  $R^2=0.67\%$ ,  $p=3.0 \times 10^{-19}$ ) was much greater than for later onset (O6-8) cases (BPD  $R^2=0.16\%$ ,  $p=1.9 \times 10^{-5}$ ; SCZ  $R^2=0.14\%$ ,  $p=3.9 \times 10^{-5}$ ; **Figure 3**). A similar increase in association with late-onset co-morbid disorders was not seen. Polygenic risk scores for coronary artery disease from the CARDIoGRAM consortium (49) were weakly positively associated with MDD, but this was consistent across early and late onset cases, in contrast to BPD and SCZ (Coronary artery disease: O1-3 vs. controls,  $R^2=0.05\%$ ,  $p=0.01$ ; O6-8 vs. controls,  $R^2=0.05\%$ ,  $p=0.01$ ; O1-3 vs. O6-8,  $R^2<0.01\%$ ,  $p=0.76$ ; **Figure 3**). No association with MDD status was seen for the scores generated for Alzheimer's disease from the GERAD

consortium (48) regardless of AAO (O1-3 vs. controls,  $R^2 < 0.01\%$ ,  $p = 0.868$ ; O6-8 vs. controls,  $R^2 = 0.02\%$ ,  $p = 0.223$ ; O1-3 vs. O6-8,  $R^2 = 0.03\%$ ,  $p = 0.157$ ; *Figure 3*). Full results from these analyses are available in ***Supplementary Table 2***.

## Discussion

Our analysis of age at onset and the genetic architecture of MDD suggested that AAO specific genetic risk factors exist. Unexpectedly, the strongest associations we observed were with the oldest half of MDD cases, where previous analyses have focused on early-onset MDD, supported by studies of the genetic epidemiology of MDD. We emphasize that here ‘late’ onset MDD was at still a relatively young age (median onset of MDD was at age 27 with 98% of our sample having onset before 60) rather than onset in old age. It is clearer to conceptualize later-onset cases here as ‘adult onset’ cases, and we recommend that this sub-group should be included in recruitment of future genetic studies, especially as no difference in the heritability captured by SNPs was observed for this subtype compared to early onset cases (see ***Supplementary Table 3***). However, substantial differences in AAO across studies make it difficult to put a precise age-cut off for this recommended ‘adult onset’ MDD.

We show significant association with rs7647854 on chromosome 3, which was associated with the 50% latest-onset cases ( $p = 5.2 \times 10^{-11}$ , OR=1.16). This SNP was identified as the second strongest association in the primary mega-analysis of the PGC-MDD data set (9), though at a much lower significance and did not replicate (discovery  $p = 6.5 \times 10^{-7}$ , replication  $p = 0.67$ ). Stratifying by AAO leads to much stronger associations in both the discovery and replication samples, despite the reduced sample size. Although analyzing across multiple overlapping AAO subtypes runs the risk of over-fitting to maximize significance, we observed highly significant associations for this SNP in both discovery and replication sample. Further we also observed a gradual increase in effect size as both raw AAO within studies and AAO percentile across studies increased (***Figure 2 and Supplementary Figure 4***). That this was seen both for raw AAO within studies and as a percentile, and in both European and Chinese ancestry studies across a wide variety of AAO measurement tools suggests that while our cut-off of the 50% latest onset cases is arbitrary, the effect exists regardless of exact cut-off used. However this does not rule out the possibility that

AAO may be a proxy for a more homogenous subgroup of MDD based on another factor (e.g. an age-specific environmental trigger or distinct pattern of symptoms with onset later in life). That median AAO and effect size across studies was not significantly correlated ( $r=0.011$ ,  $p=0.68$ ) despite greater effect sizes by AAO percentile within studies, suggests that differences in measurement of AAO across studies might obscure effects. rs7647854 is intergenic, with flanking genes including *C3orf70*, *VPS8*, *EHHADH* and *MAP3K13*. *C3orf70*, *VPS8*, and *MAP3K13* all show evidence of expression in the brain in several areas of potential interest for MDD including the hypothalamus, frontal cortex, pituitary and thyroid (GTex; <http://www.gtexportal.org/>) (53). *VPS8* and *MAP3K13* also show a slight increase in (RNA) expression within various brain regions during neo-natal development (Human Brain Transcriptome; <http://hbatlas.org/>) (54) with maintained expression into adulthood.

This genetic association arises with the supposedly less heritable form of adult-onset MDD, although such summary measures give no information on the effect sizes of individual SNPs. One potential explanation is greater contamination of early-onset MDD cases by individuals misclassified with MDD or having co-morbid disorders. Longitudinal studies show that early-onset depressive symptoms predict not just adult depression but also psychosis (e.g. 55), and there is significant genetic overlap between MDD and other psychiatric disorders (50). Inclusion of early-onset cases with individuals who will later develop schizophrenia or bipolar disorder would reduce the power of genome-wide association studies for MDD, though secondary analysis of age at interview did not support this (**Supplementary Figure 7**). The polygenic risk score results show that early and adult-onset MDD cases differ in their genetic susceptibility to bipolar disorder and schizophrenia and suggest that some of the heterogeneity in MDD results from the inclusion of early onset cases with a greater genetic overlap or misclassification with these disorders. We did not observe a similar pattern of association for two disorders often co-morbid with late-onset MDD, Alzheimer's disease and coronary artery disease (51; 52). We did, however, show for the first time a genetic overlap between MDD and coronary artery disease irrespective of AAO but not for Alzheimer's disease. We also found that the heritability explained by SNPs across the genome for early and late-onset MDD did not significantly differ suggesting, for common variants at least, recruiting only early-onset cases would not increase power (**Supplementary Table 3**).



The limitations of our approach are the reduced sample size and multiple testing from stratifying cases into subtypes, and requiring potentially less reliable secondary phenotypes to be widely collected. Measures such as AAO rely on self-report, and are often assessed differently across studies that can be problematic for comparison. The analyses presented here addressed this by analyzing AAO relative to the median of a study, though this assumed each study recruited cases from the same distribution of onset with observed differences due to how AAO was defined. Further, the effect of our genome-wide associated SNP were consistent in the replication studies, ethnicities and countries, suggesting that differences in measurement may not be as much of a limitation as expected, at least in the case of AAO in MDD. The other disadvantage of looking at more homogenous groups is the reduction in sample size. However, it has previously been shown that only modest increases in effect size may be required to offset the reduction in power from analyzing fewer cases, implying that analyses of more homogenous subgroups have the potential to identify novel associations (56).

Our study illustrates the value of using additional phenotypic information on cases in genome-wide association studies. We show here that including information on AAO increases the power to detect associations with MDD, and that analyzing polygenic risk scores from related diseases enables us to identify sources of phenotypic heterogeneity that may have hampered previous genetic studies. In contrast to other approaches that weight cases on AAO (57), our stratification of cases on AAO is agnostic to the direction of the phenotypic effect. Our approach uses the additional phenotype data previously collected by genetic studies and complements an alternative emphasis on collecting large sample sizes through minimal phenotyping. Both strategies will undoubtedly be necessary to identify and characterize different components of the genetic architecture of psychiatric disorders. Further, our analysis shows that tackling phenotypic heterogeneity in MDD can successfully lead to the discovery of genetic risk factor despite reduced sample size. The identification here of a novel genetic risk variant for MDD is of great importance due to both the scarcity of evidence for the underlying biology and its pressing economic health burden.



## Acknowledgments:

The PGC was funded by NIMH Grants MH085520 (lead PI PF Sullivan) and MH080403. Statistical analyses were carried out on the Genetic Cluster Computer (<http://www.geneticcluster.org>) hosted by SURFsara and financially supported by the Netherlands Scientific Organization (NWO 480-05-003 PI: Posthuma) and the department of Psychology, Vrije Universiteit Amsterdam along with a supplement from the Dutch Brain Foundation. The Bonn/Mannheim (BoMa) GWAS was supported by the German Federal Ministry of Education and Research (BMBF) through the Integrated Genome Research Network (IG) MooDS (Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia; grants 01GS08144 and 01GS08147), under the auspices of the National Genome Research Network plus (NGFNplus), and through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders), under the auspices of the e:Med Programme (grants 01ZX1314A and 01ZX1314G). The BoMa GWAS was also supported by the German Research Foundation (DFG; grant FOR2107; RI908/11-1 and NO246/10-1). The GenRED GWAS project was supported by NIMH R01 Grants MH061686 (DF Levinson), MH059542 (WH Coryell), MH075131 (WB Lawson), MH059552 (JB Potash), MH059541 (WA Scheftner) and MH060912 (MM Weissman). We acknowledge the contributions of Dr George S Zubenko and Dr Wendy N Zubenko, Department of Psychiatry, University of Pittsburgh School of Medicine, to the GenRED I project. The NIMH Cell Repository at Rutgers University and the NIMH Center for Collaborative Genetic Studies on Mental Disorders made essential contributions to this project. Genotyping was carried out by the Broad Institute Center for Genotyping and Analysis with support from Grant U54 RR020278 (which partially subsidized the genotyping of the GenRED cases). Collection and quality control analyses of the control data set were supported by grants from NIMH and the National Alliance for Research on Schizophrenia and Depression. We are grateful to Knowledge Networks (Menlo Park, CA, USA) for assistance in collecting the control data set. We express our profound appreciation to the families who participated in this project, and to the many clinicians who facilitated the referral of participants to the study. Max Planck Institute of Psychiatry MARS study was supported by the BMBF Program Molecular Diagnostics: Validation of Biomarkers for Diagnosis and Outcome in Major Depression (01ES0811). Genotyping was supported by the Bavarian Ministry of Commerce, and the Federal Ministry of Education and Research (BMBF) in the framework of the National Genome Research Network (NGFN2 and NGFN-Plus, FKZ

01GS0481 and 01GS08145). The Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Twin Register (NTR) contributed to GAIN-MDD and to MDD2000. Funding was from: the Netherlands Organization for Scientific Research (MagW/ZonMW Grants 904-61-090, 985-10- 002, 904-61-193, 480-04-004, 400-05-717, 912-100-20; Spinozapremie 56-464-14192; Geestkracht program Grant 10-000-1002); the Center for Medical Systems Biology (NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure, VU University's Institutes for Health and Care Research and Neuroscience Campus Amsterdam, BIC/BioAssist/RK (2008.024); the European Science Foundation (EU/QLRT-2001-01254); the European Community's Seventh Framework Program (FP7/2007-2013); ENGAGE (HEALTH-F4-2007-201413); and the European Science Council (ERC, 230374). Genotyping was funded in part by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). CM Middeldorp was supported by the Netherlands Organization for Scientific Research (NOW VENI grant 916-76-125). Funding for the QIMR samples was provided by the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496675, 496739, 552485, 552498, 613602, 613608, 613674, 619667), the Australian Research Council (FT0991360, FT0991022), the FP-5 GenomeUtwinn Project (QLG2-CT- 2002-01254) and the US National Institutes of Health (AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, MH66206, DA12854, DA019951), and the Center for Inherited Disease Research (Baltimore, MD, USA). We thank the twins and their families registered at the Australian Twin Registry for their participation in the many studies that have contributed to this research. RADIANT was funded by: a joint grant from the UK Medical Research Council and GlaxoSmithKline (G0701420); the National Institute for Health Research (NIHR) Specialist Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and the Institute of Psychiatry, King's College London; the UK Medical Research Council (G0000647), and the Marie Curie Industry-Academia Partnership and Pathways, grant agreement n° 286213. The GENDEP study was funded by a European Commission Framework 6 grant, EC Contract Ref.: LSHB-CT-2003-503428. Genotyping of STAR\*D was supported by an NIMH Grant MH072802 (SP Hamilton). STAR\*D was funded by the National Institute of Mental Health (contract N01MH90003) to the University of Texas Southwestern Medical Center at Dallas (AJ Rush, principal investigator). The CoLaus/PsyCoLaus study was supported by research grants from GlaxoSmithKline, the Faculty of Biology and

Medicine of Lausanne, and the Swiss National Science Foundation (grants 3200B0–105993, 3200B0-118308, 33CSCO-122661, 33CS30-139468 and 33CS30-148401) and two grants from GlaxoSmithKline Clinical Genetics. We also thank V Mooser, G Weaber and P Vollenweider who initiated the CoLaus project. We express our gratitude to the Lausanne inhabitants who volunteered to participate in the PsyCoLaus study. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (Grants no. 01ZZ9603, 01ZZ0103 and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data have been supported by the Federal Ministry of Education and Research (Grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania. SHIP-LEGEND is funded by the German Research Foundation (DFG: GR 1912/5-1). The TwinGene study was supported by the Swedish Ministry for Higher Education, the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT- 2001-01254; QLG2-CT-2002-01254), the Swedish Foundation for Strategic Research and the US National Institutes of Health (U01 DK066134). We would like to acknowledge the Prisme-study group, Denmark for collection of the PRISME samples. The collection of controls and genotyping of the 883 Danish controls was supported by grants from The Danish Strategic Research Council, The Stanley Research Foundation and H. Lundbeck A/S. We thank David M Hougaard, Section of Neonatal Screening and Hormones, Statens Serum Institute, Copenhagen, Denmark, Preben Bo Mortensen, National Centre for Register-based Research, Aarhus University, Denmark, Merete Nordentoft, Mental Health Centre Copenhagen, Denmark and The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark. The Muenster Depression cohorts were supported by the EU (N Health-F2-2008-222963) and by grants from the German Research Foundation (DFG; grant FOR 2107; DA1151/5-1 to UD, Innovative Medizinische Forschung (IMF) of the Medical Faculty of Münster (DA120903 to UD, DA111107 to UD, and DA211012 to UD). Generation Scotland is supported by a Wellcome Trust Strategic Award “Stratifying Resilience and Depression Longitudinally” (STRaDL) (Reference 104036/Z/14/Z) and core support from the Chief Scientist Office (CSO) of the Scottish Government Health Directorates [grant number CZD/16/6] and the Scottish Funding Council [HR03006]. DJM is supported by an NRS Career Fellowship, funded by the CSO. AMM is supported by a Scottish Funding Council Senior Clinical Fellowship and by the Dame Theresa and Mortimer Sackler Foundation. CSH acknowledges support from the Medical

Research Council (MRC) and the Biotechnology and Biological Sciences Research Council (BBSRC). IJD is supported by the MRC-BBSRC, Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (MR/K026992/1). Funding from the BBSRC and MRC is gratefully acknowledged.

**Conflict of Interest:**

Volker Arolt received funds from the German Federal Ministry of Education and Research, from the European Union (FP 7) and from the Interdisciplinary Center for Clinical Research Münster. He has been a member of the advisory boards of, or has given presentations on behalf of the following companies: Astra-Zeneca, Janssen-Organon, Lilly, Lundbeck, Servier, Otsuka, and Trommsdorff. Volker Arolt is a member of the advisory board of, or has given presentations on behalf of, the following companies: Astra-Zeneca, Janssen-Organon, Lilly, Lundbeck, Servier, Pfizer, Otsuka, and Trommsdorff. Bernhard T Baune receives funding from the National Health and Medical Research Council Australia and honoraria from Lundbeck, Bristol-Meyers Squibb, Sanofi, Servier, Astra-Zeneca, Pfizer. Andrew McIntosh has received research support from Pfizer, Janssen and Lilly. Hans J Grabe reports receiving funding from: German Research Foundation; Federal Ministry of Education and Research Germany; speakers honoraria from, Eli Lilly, and Servier. Bertram Muller-Myhsok consulted for Affectis Pharmaceuticals. Martin Preisig has been member of the advisory boards of Lundbeck and Eli Lilly. Stacy Steinberg, Kari Stefansson, Hreinn Stefansson and Thorgeir E. Thorgeirsson are employees of deCODE Genetics/Amgen.

No other author reports a conflict of interest.

## References

1. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, *et al.* (2003): The epidemiology of major depressive disorder - Results from the National Comorbidity Survey Replication (NCS-R). *Jama-J Am Med Assoc* 289: 3095-3105.
  2. Mueller TI, Leon AC, Keller MB, Solomon DA, Endicott J, Coryell W, *et al.* (1999): Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. *The American journal of psychiatry* 156: 1000-1006.
  3. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, *et al.* (2012): Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380: 2163-2196.
- American Psychiatric Association, 1994. Diagnostic and Statistical Manual of Mental Disorders, 4th edn. American Psychiatric Association, Washington, DC.
5. Angst F, Stassen HH, Clayton PJ, Angst J. (2002): Mortality of patients with mood disorders: follow-up over 34-38 years. *Journal of affective disorders* 68: 167-181.
  6. Barth J, Schumacher M, Herrmann-Lingen C. (2004): Depression as a risk factor for mortality in patients with coronary heart disease: a meta-analysis. *Psychosomatic medicine* 66: 802-813.
  7. Satin JR, Linden W, Phillips MJ. (2009): Depression as a predictor of disease progression and mortality in cancer patients: a meta-analysis. *Cancer* 115: 5349-5361.
  8. Sullivan PF, Neale MC, Kendler KS. (2000): Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry* 157: 1552-1562.
  9. Major Depressive Disorder Working Group of the Psychiatric Gwas Consortium, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, *et al.* (2013): A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry* 18: 497-511.
  10. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, *et al.* (2011): Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 43: 969-U977.
  11. Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, *et al.* (2011): Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 43: 977-983.
  12. Gillum LA, Gouveia C, Dorsey ER, Pletcher M, Mathers CD, Mcculloch CE, *et al.* (2011): NIH disease funding levels and burden of disease. *PloS one* 6: e16837.
  13. Sirey JA, Bruce ML, Alexopoulos GS, Perlick DA, Raue P, Friedman SJ, *et al.* (2001): Perceived stigma as a predictor of treatment discontinuation in young and older outpatients with depression. *The American journal of psychiatry* 158: 479-481.
  14. Weissman MM, Wickramaratne P, Merikangas KR, Leckman JF, Prusoff BA, Caruso KA, *et al.* (1984): Onset of major depression in early adulthood. Increased familial loading and specificity. *Archives of general psychiatry* 41: 1136-1143.
  15. Nierenberg AA, Trivedi MH, Fava M, Biggs MM, Shores-Wilson K, Wisniewski SR, *et al.* (2007): Family history of mood disorder and characteristics of major depressive disorder: a STAR\*D (sequenced treatment alternatives to relieve depression) study. *J Psychiatr Res* 41: 214-221.
  16. Tozzi F, Prokopenko I, Perry JD, Kennedy JL, McCarthy AD, Holsboer F, *et al.* (2008): Family history of depression is associated with younger age of onset in patients with recurrent depression. *Psychol Med* 38: 641-649.
  17. Lyons MJ, Eisen SA, Goldberg J, True W, Lin N, Meyer JM, *et al.* (1998): A registry-based twin study of depression in men. *Archives of General Psychiatry* 55: 468-472.
  18. Yang F, Li Y, Xie D, Shao C, Ren J, Wu W, *et al.* (2011): Age at onset of major depressive disorder in Han Chinese women: relationship with clinical features and family history. *J Affect Disord* 135: 89-94.
  19. Wickramaratne PJ, Weissman MM. (1998): Onset of psychopathology in offspring by developmental phase and parental depression. *J.Am.Acad.Child Adolesc.Psychiatry* 37: 933-942.
  20. Kendler KS, Gatz M, Gardner CO, Pedersen NL. (2005): Age at onset and familial risk for major depression in a Swedish national twin sample. *Psychol Med* 35: 1573-1579.
  21. Gillespie NA, Kirk KM, Evans DM, Heath AC, Hickie IB, Martin NG. (2004): Do the genetic or environmental determinants of anxiety and depression change with age? A longitudinal study of Australian twins. *Twin research : the official journal of the International Society for Twin Studies* 7: 39-53.



22. Nivard MG, Dolan CV, Kendler KS, Kan KJ, Willemsen G, Van Beijsterveldt CE, *et al.* (2015): Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. *Psychological medicine* 45: 1039-1049.
23. Rietschel M, Mattheisen M, Frank J, Treutlein J, Degenhardt F, Breuer R, *et al.* (2010): Genome-wide association-, replication-, and neuroimaging study implicates HOMER1 in the etiology of major depression. *Biological psychiatry* 68: 578-585.
24. Sullivan PF, De Geus EJC, Willemsen G, James MR, Smit JH, Zandbelt T, *et al.* (2009): Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Molecular Psychiatry* 14: 359-375.
25. Shi J, Potash JB, Knowles JA, Weissman MM, Coryell W, Scheftner WA, *et al.* (2011): Genome-wide association study of recurrent early-onset major depressive disorder. *Molecular Psychiatry* 16: 193-201.
26. Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ, *et al.* (2010): Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Molecular Psychiatry* 15: 589-601.
27. Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR, *et al.* (2010): Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry*.
28. Shyn SI, Shi J, Kraft JB, Potash JB, Knowles JA, Weissman MM, *et al.* (2011): Novel loci for major depression identified by genome-wide association study of Sequenced Treatment Alternatives to Relieve Depression and meta-analysis of three studies. *Mol Psychiatry* 16: 202-215.
29. Lewis CM, Ng MY, Butler AW, Cohen-Woods S, Uher R, Pirlo K, *et al.* (2010): Genome-Wide Association Study of Major Recurrent Depression in the UK Population. *American Journal of Psychiatry* 167: 949-957.
30. Ising M, Lucae S, Binder EB, Bettecken T, Uhr M, Ripke S, *et al.* (2009): A Genomewide Association Study Points to Multiple Loci That Predict Antidepressant Drug Treatment Outcome in Depression. *Archives of General Psychiatry* 66: 966-+.
31. Browning BL, Browning SR. (2009): A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet* 84: 210-223.
32. International Hapmap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, *et al.* (2010): Integrating common and rare genetic variation in diverse human populations. *Nature* 467: 52-58.
33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MaR, Bender D, *et al.* (2007): PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575.
34. Macgregor S, Craddock N, Holmans PA. (2006): Use of phenotypic covariates in association analysis by sequential addition of cases. *Eur J Hum Genet* 14: 529-534.
35. Lichtenstein P, Sullivan PF, Cnattingius S, Gatz M, Johansson S, Carlstrom E, *et al.* (2006): The Swedish Twin Registry in the third millennium: an update. *Twin research and human genetics : the official journal of the International Society for Twin Studies* 9: 875-882.
36. Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandeleur C, *et al.* (2009): The PsyCoLaus study: methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. *BMC psychiatry* 9: 9.
37. Volzke H, Alte D, Schmidt CO, Radke D, Lohrer R, Friedrich N, *et al.* (2011): Cohort profile: the study of health in Pomerania. *International journal of epidemiology* 40: 294-307.
38. Baune BT, Konrad C, Grotegerd D, Suslow T, Ohrmann P, Bauer J, *et al.* (2012): Tumor necrosis factor gene variation predicts hippocampus volume in healthy individuals. *Biological psychiatry* 72: 655-662.
39. Grosse L, Carvalho LA, Wijkhuijs AJ, Bellingrath S, Ruland T, Ambree O, *et al.* (2015): Clinical characteristics of inflammation-associated depression: Monocyte gene expression is age-related in major depressive disorder. *Brain, behavior, and immunity* 44: 48-56.
40. Uher R, Perroud N, Ng MYM, Hauser J, Henigsberg N, Maier W, *et al.* (2010): Genome-Wide Pharmacogenetics of Antidepressant Response in the GENDEP Project. *American Journal of Psychiatry* 167: 555-564.
41. Krogh J, Petersen L, Timmermann M, Saltin B, Nordentoft M. (2007): Design paper: the DEMO trial: a randomized, parallel-group, observer-blinded clinical trial of aerobic versus non-aerobic versus relaxation training for patients with light to moderate depression. *Contemporary clinical trials* 28: 79-89.

42. Elfving B, Buttenschon HN, Foldager L, Poulsen PH, Andersen JH, Grynderup MB, *et al.* (2012): Depression, the Val66Met polymorphism, age, and gender influence the serum BDNF level. *Journal of psychiatric research* 46: 1118-1125.
43. Yang F, Zhao H, Wang Z, Tao D, Xiao X, Niu Q, *et al.* (2014): Age at onset of recurrent major depression in Han Chinese women - a replication study. *Journal of affective disorders* 157: 72-79.
44. Consortium C. (2015): Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* 523: 588-591.
45. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, *et al.* (2013): Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* 42: 689-700.
46. Willer CJ, Li Y, Abecasis GR. (2010): METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26: 2190-2191.
47. Purcell SM, Wray NR, Stone JL, Visscher PM, O'donovan MC, Sullivan PF, *et al.* (2009): Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 460: 748-752.
48. Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML, *et al.* (2009): Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 41: 1088-1093.
49. Schunkert H, König IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, *et al.* (2011): Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 43: 333-338.
50. Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, Nurnberger JI, *et al.* (2013): Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* 381: 1371-1379.
51. Taylor WD, Aizenstein HJ, Alexopoulos GS. (2013): The vascular depression hypothesis: mechanisms linking vascular disease with depression. *Mol Psychiatry* 18: 963-974.
52. Starkstein SE, Jorge R, Mizrahi R, Robinson RG. (2005): The construct of minor and major depression in Alzheimer's disease. *The American journal of psychiatry* 162: 2086-2093.
53. Consortium GT. (2013): The Genotype-Tissue Expression (GTEx) project. *Nature genetics* 45: 580-585.
54. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, *et al.* (2011): Spatio-temporal transcriptome of the human brain. *Nature* 478: 483-489.
55. Kim-Cohen J, Caspi A, Moffitt TE, Harrington H, Milne BJ, Poulton R. (2003): Prior juvenile diagnoses in adults with mental disorder - Developmental follow-back of a prospective-longitudinal cohort. *Archives of General Psychiatry* 60: 709-717.
56. Traylor M, Markus H, Lewis CM. (2014): Homogeneous case subgroups increase power in genetic association studies. *European journal of human genetics : EJHG*.
57. Zaitlen N, Lindstrom S, Pasaniuc B, Cornelis M, Genovese G, Pollack S, *et al.* (2012): Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS genetics* 8: e1003032.

**Table 1:**  
**Summary of MDD discovery and replication cohorts. AAO stands for age at onset.**

	Study	Country	Measure of AAO	Cases with AAO	Controls	Median AAO
Discovery	NESDA/NTR (GAIN)	The Netherlands	CIDI	1,675	1,765	25
	GenRED	US	DIGS3	1,020	1,253	16
	GSK	Germany	SCAN	887	864	36
	MDD2000-QIMR_610	AU	CIDI/SSAGA	432	751	26
	MDD2000-QIMR_317	AU	CIDI/SSAGA	1,015	960	26
	MPIP	Germany	Asked at interview	373	537	37
	RADIANT Bonn/Mannheim	Germany	SCAN	883	1,290	33
	RADIANT	UK	SCAN	1,407	1,588	20
	STAR*D	US	Asked at interview	1,228	511	21
	Total			8,920	9,519	
Replication	TwinGene	Sweden	SALT	1,009	8,601	40
	PsyCoLaus	Switzerland	DIGS	1,358	1,687	33
	SHIP-LEGEND	Germany	M-CIDI	381	1,827	37
	GenRED2/DepGenesNetworks	US	DIGS3	1,296	930	17
	University of Münster	Germany	SCID	402	516	27
	Combined Danish sample	Denmark	SCAN	461	1,197	31

	CONVERGE	China		5,715	5,537	34
	deCODE	Iceland	Hospital records	1,005	99,175	39
	Generation Scotland	UK	SCID	1,611	4,760	30
	Total			13,238	124,230	
Total				22,158	133,749	

**Table 2**

**Summary of association with rs7647854, located at 3q27.2 (186359477 base pairs), with odds ratios (OR) for the reference allele, G (frequency 0.16) compared to the non-reference allele A.** All results reported are for the oldest half of MDD cases (O5-8), which had the strongest evidence for association in the discovery study.

Study	No. of MDD cases	Number of Controls	OR (95% CI)	P-value
Discovery	3,869	9,519	1.30 (1.20-1.40)	$3.4 \times 10^{-11}$
Replication	6,107	124,230	1.10 (1.05- 1.17)	$7.5 \times 10^{-4}$
Meta-analysis	9,976	133,749	1.16 (1.11-1.21)	$5.2 \times 10^{-11}$

**Figure Legends:****Figure 1:**

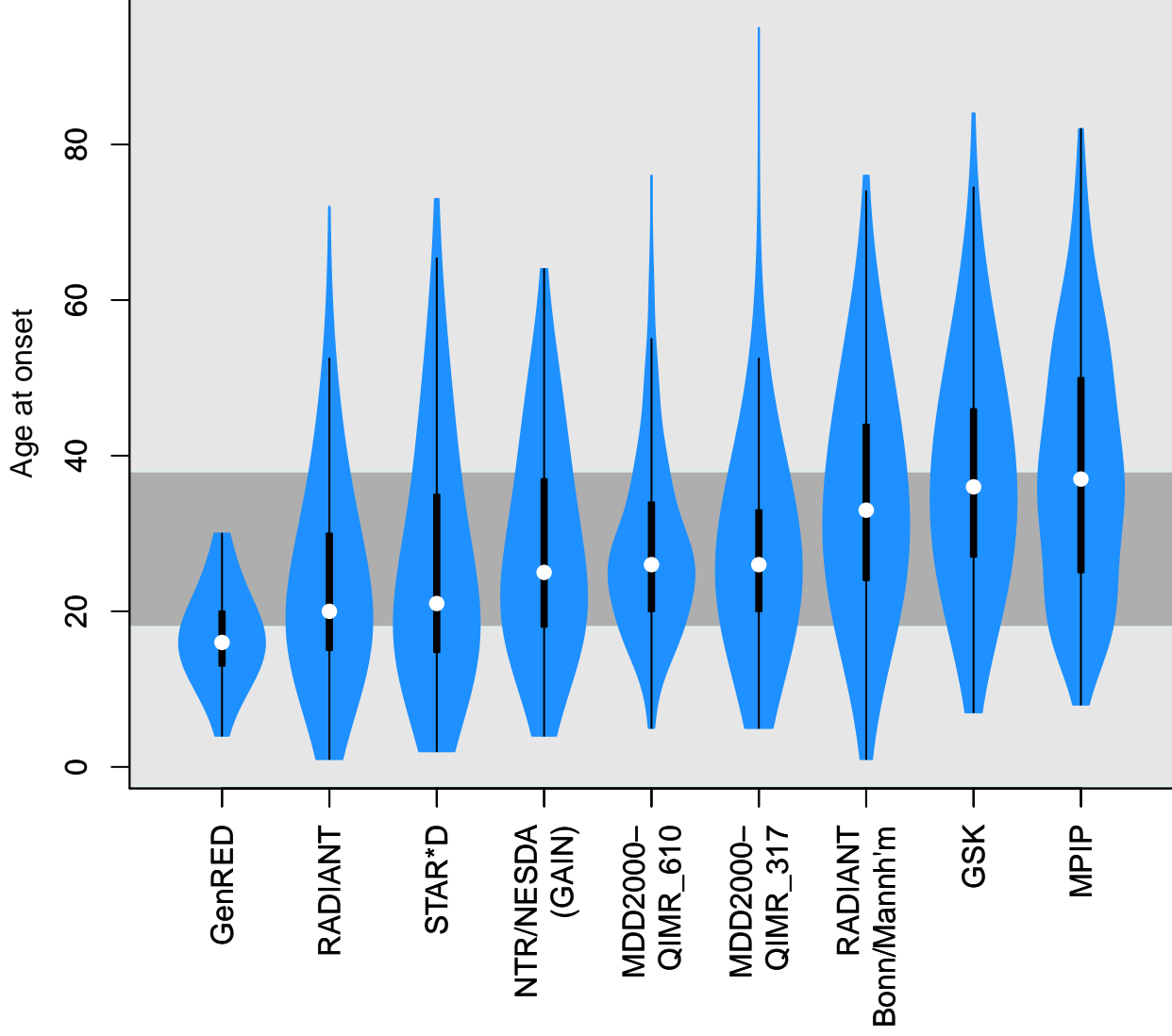
**Distribution of age at onset across the nine studies included in the discovery analysis.** Mid-grey band shows interquartile range across all studies excluding GENDEP, which recruited only cases onset at 30 years or below.

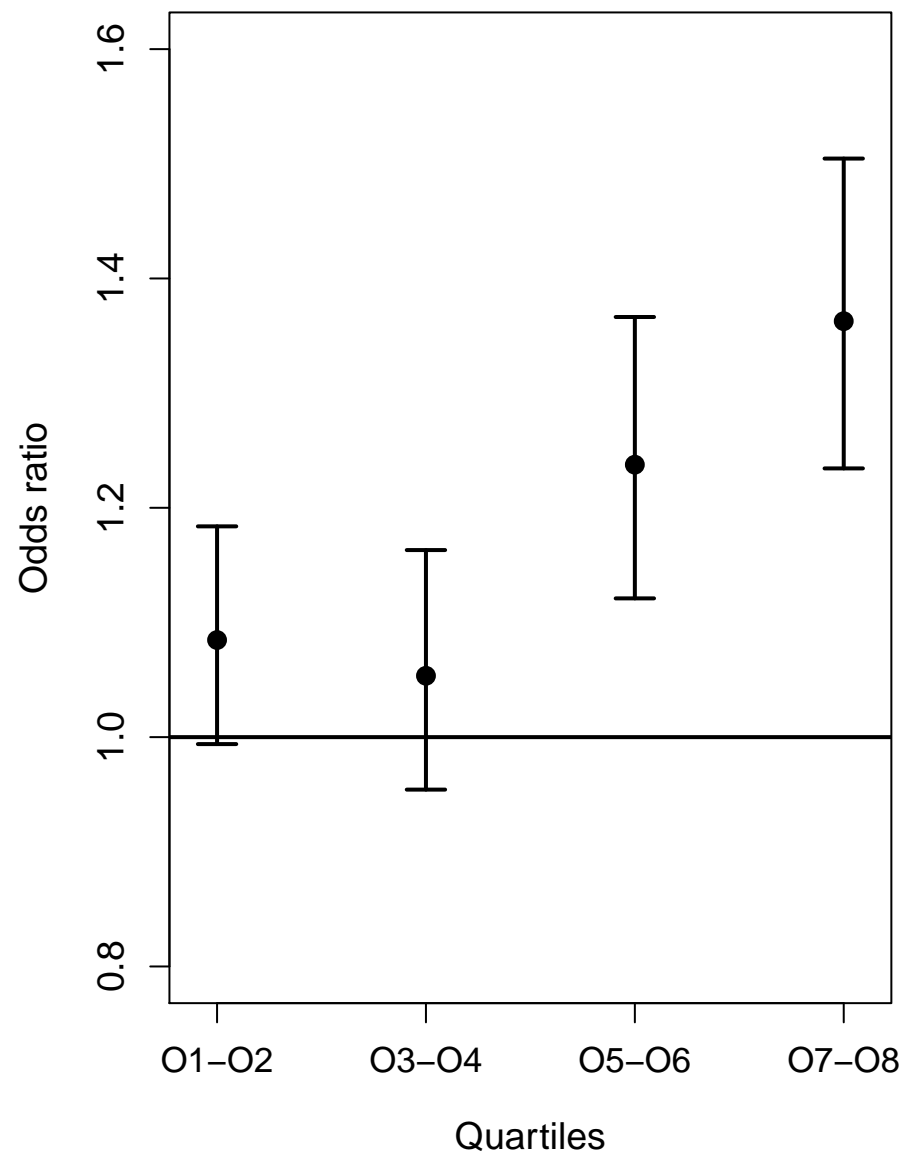
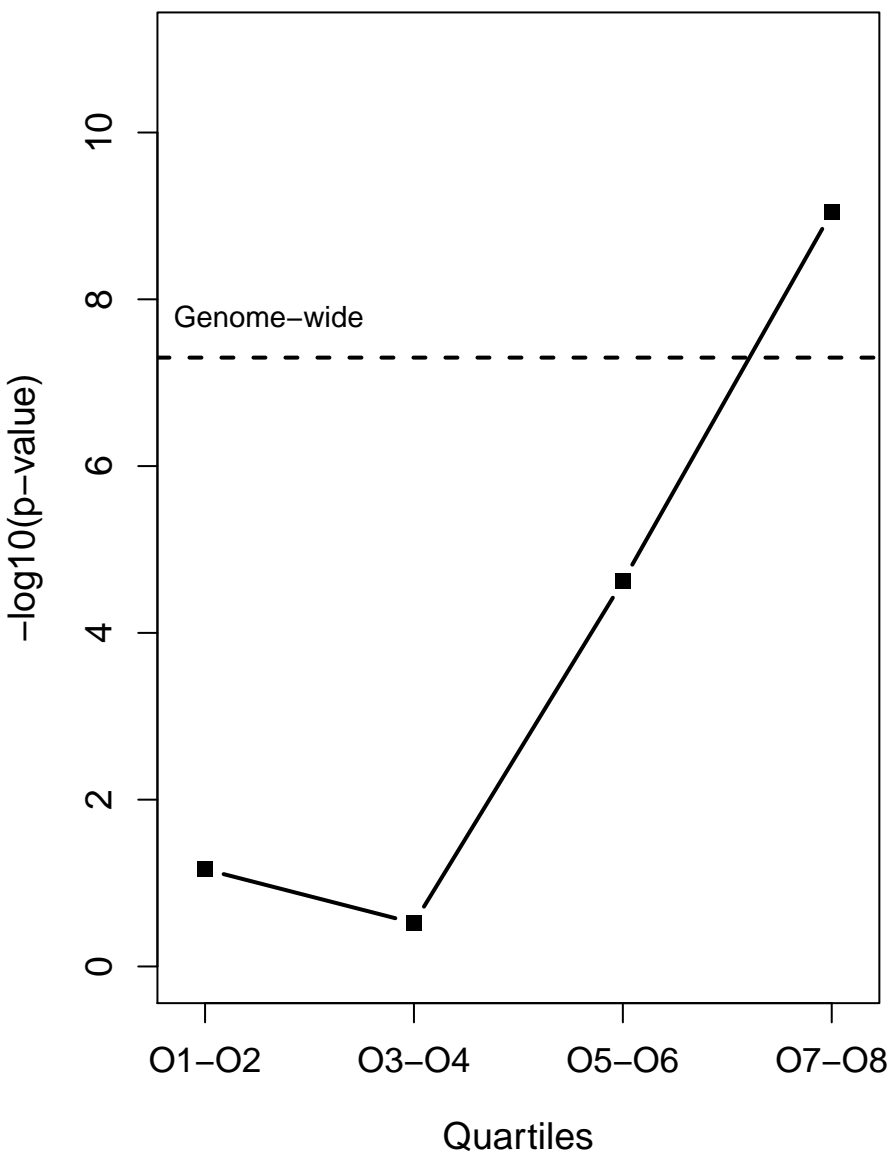
**Figure 2:**

**Evidence for association and effect size for rs7647854 on chromosome 3, with cases split into non-overlapping quartiles by age at onset within discovery studies.**

**Figure 3:**

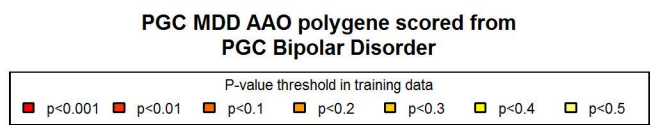
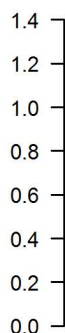
**Polygenic risk profile scoring analysis of bipolar disorder, schizophrenia, Alzheimer's disease and coronary artery disease within the MDD discovery studies (excluding GenRED).** We calculated the proportion of variance explained (Nagelkerke's  $R^2$ ) by subtraction of a full model (covariates + RPS) score from a reduced model (covariates only).







R-squared (%)



All Cases v Controls

O1-O3 v Controls

O6-O8 v Controls

O1-O3 v O6-O8

PGC Schizophrenia

All Cases v Controls

O1-O3 v Controls

O6-O8 v Controls

O1-O3 v O6-O8

Alzheimer's Disorder

All Cases v Controls

O1-O3 v Controls

O6-O8 v Controls

O1-O3 v O6-O8

Coronary Artery Disease

All Cases v Controls

O1-O3 v Controls

O6-O8 v Controls

O1-O3 v O6-O8